DOCUMENT RESUME

ED 329 582 TM 016 213

AUTHOR Shannon, David M.
TITLE Teacher Evaluation: A Functional Approach.
PUB DATE Feb 91
NOTE 16p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (14th, Boston, MA, February 13-17, 1991).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Assisted Testing; Computer Simulation; Elementary School Teachers; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; Higher Education; *Job Performance; *Occupational Tests; Secondary School Teachers; Standardized Tests; *Teacher Evaluation; Test Construction; Test Format; *Test Use; Test Validity
IDENTIFIERS Paper and Pencil Tests; *Simulation Test Interact Teach Competencies (Hays); Subject Content Knowledge; *Teacher Competency Testing

ABSTRACT

An overview is presented of methods of teacher assessment. The most heavily used methods of teacher assessment are standardized paper-and-pencil examinations and on-the-job ratings by supervisors. An alternative to the paper-and-pencil approach is explored, and preliminary evidence is presented in support of this assessment strategy. Paper-and-pencil tests have been criticized primarily for their lack of face validity and predictive validity. Ratings by supervisors have been criticized for: (1) lack of accurate measurement properties; (2) low predictive validity for student learning; and (3) halo effects. An alternative, a modification of traditional paper-and-pencil tests, is the Simulation Test of Interactive Teaching Competencies-Hays (STITC-H), developed by L. Hays. Four alternative courses of teacher action are presented for choice in the STITC-H using a slide projector and an audiocassette. Preliminary evidence from a study by Hays involving 46 experienced teachers, 30 student teachers, and 31 adults without teacher training suggested the usefulness of the test as a measure of teacher learning or experience. D. M. Shannon (1990) revised the STITC-H and investigated it with 25 student teachers and 78 non-education college students. Evidence of construct validity was obtained, and this alternative approach to testing teachers was supported. A 23-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED329582

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

DAVID M. SHANNON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Teacher Evaluation: A Functional Approach

David M. Shannon

Auburn University
2084 Haley Center
Auburn, AL 36849
(205) 844-3071

Paper Presented at the Annual Meeting of the
Eastern Educational Research Association (EERA),
Boston, MA.
February 13-17, 1991

RUNNING HEAD: Teacher Evaluation

TM016213

Abstract

This paper has three primary purposes. The first section is an overview of the two most-heavily used methods of teacher assessment: standardized pencil-and-paper examinations and on the job supervisory ratings. This overview includes a discussion of the advantages and disadvantages associated with each method as well the general support for and expansion of the assessment of the competence of teachers. The emphasis is placed on teacher tests most frequently mandated by state legislatures. Secondly, alternatives to such paper-and-pencil tests are explored and one such alternative is suggested. Finally, preliminary evidence in support of this alternative assessment strategy is presented.

Teacher Evaluation: A Functional Approach

Teacher evaluation has been used as a gauge of the competence of teachers in our schools. The decline of students' test scores below the public's expectations have led the public to believe that part of this problem can be linked to teacher preparation. Therefore, in response to public opinion, state legislatures and state departments of education have mandated that teachers be evaluated to assure their competence in the classroom. The implementation of such evaluation systems is expected to result in more effective teaching practices and, in turn, return the level of student achievement to an acceptable level in the eye of the public. Traditionally, the evaluation of teachers has been conducted with standardized paper and pencil tests or classroom performance ratings by a supervisor.

Performance Ratings

Performance ratings by supervisors have been reported as the most common method used to evaluate teachers (Medley, 1982: Scriven, 1981). While these ratings have been traditionally used for the evaluation of both student teachers and experienced teachers, fifteen states now mandate performance assessments before certification (Association of Teacher Educators, 1988).

In addition, Educational Testing Service (ETS) has proposed a teacher assessment project for the primary purpose of state licensure which outlines the assessment of preservice teachers at three different stages and is scheduled to replace the existing

National Teacher Examination by 1992 (Givens, 1988). The last stage of assessment would occur after the student has had classroom experience and includes performance measures obtained from computer simulations, classroom observations, and other evidence collected during student teaching, internships, or other supervised experiences. Both the National Education Association (Hodgkins & McKenna, 1982) and others (Piper & Houston, 1981) have recommended performance measures over paper and pencil exams indicating that the best predictor of successful teaching is success in teaching situations.

Despite their widespread use, ratings by supervisors have been criticized for three primary reasons. These criticisms have been aimed at the lack of accurate measurement properties, low predictive validity of student learning and "halo" effects (Medley, Coker & Soar, 1984). First, rating scales lack the minimum properties necessary for accurate measurement: 1) a standard task, 2) a record of performance, 3) an agreed-upon scoring key, and 4) publicly available norms. Second, they have little or no predictive validity of teacher effectiveness as measured by student learning. Third, rating scales allow raters the opportunity to express how competent they feel the teacher is, which is determined by the overall impression the teacher makes on the rater.

Paper-and-Pencil tests

State legislatures and boards of education have generally

relied on paper and pencil tests. This has been the case because paper-and-pencil tests are relatively cheap, quick, easy to administer and readily available. These tests have been designed to include basic skills, general knowledge, subject-matter knowledge, and professional knowledge. Several different tests are used, including tests developed by national testing organizations such as Educational Testing Service (ETS) and National Evaluation Systems (NES) and State Departments of Education. Eissenburg and Rudner (1988) indicated that 46 states assess or plan to assess teachers for either admission or certification purposes.

Criticisms of Paper-and-Pencil Tests

These tests have been criticized primarily because of their lack of face validity and predictive validity. The task of choosing the one "best" answer on a multiple-choice test is much different from reacting to situations in the classroom for which there may not be a "best" alternative (Medley, 1978; Pottinger, 1979). Palladino (1980) (cited in Hays, 1988), identified another problem associated with the neutrality of multiple-choice tests. He asserted that "correct" answers should vary depending on differences in basic philosophy. Without this flexibility, standardization may strongly influence teacher training problems causing them to "teach to the test", narrowing the opportunities and viewpoints afforded to preservice teachers, conforming at the expense of creativity (Palladino, in Hays, 1988; Vold, 1985).

There is also very little empirical evidence that these tests can predict teaching effectiveness. Quirk, Witten, and Weinberg (1973) have compiled a review of the predictive validity studies conducted during the first 30 years of the NTE's existence. Quirk, et al. (1973) reported a median correlation of .05 between NTE scores and ratings by college supervisors and principals during student teaching. Chiu (1989) examined sixteen predictive validity studies which reported a total of 168 bivariate correlations of the NTE with classroom performance ratings by principals, students and with low-inference observational systems. The median correlation was .15. A review of three predictive validity studies (79 bivariate correlations) using the Core Battery from the revised NTE tests resulted in a median of .09.

Academic and Functional Knowledge of Teachers

Attempts to measure the professional knowledge of teachers through objective pencil-and-paper have been limited to the measurement of academic professional knowledge. This type of professional knowledge is available to the teacher at the time of the examination, but it is not the kind of knowledge that the teacher needs in order to function in the classroom. When a teacher is taking these tests, she knows that each problem has one correct answer. When the same teacher is teaching, problems may have one answer, more than one answer, or none.

It has been the traditional role of colleges of teacher

education to investigate, develop, maintain, and pass along professional knowledge based on research. This academic knowledge must be used selectively and when the appropriate situation arises. Unfortunately, this knowledge rarely provides for direct solutions to professional problems encountered as classroom teachers. Academic knowledge must become functional in order to be useful.

The knowledge that a teacher possesses must be functional within the classroom (Medley, Rosenblum, & Vance, 1989). This functional professional knowledge is the knowledge that a teacher draws upon when actually teaching in the classroom. The limited number of hours available for professional education courses, practicum experiences, and faculty clinical supervision leaves the teacher education programs less equipped to provide the functional knowledge necessary for effective teaching.

Smith (1980) has referred to this knowledge as clinical knowledge. The ability to apply this knowledge effectively in practice can best be learned through continued clinical experience as a teacher. This knowledge is that which effective teachers draw upon when teaching in their own classrooms. The assessment of teachers' professional knowledge should focus on functional knowledge of the principal research findings on teacher effectiveness.

Alternative to Paper-and-Pencil Tests

Despite the limitations associated with paper-and-pencil tests of the professional knowledge of teachers, the demands for such tests from state legislatures and teacher training institutions remains constant. As discussed thus far, the task of taking a multiple-choice test is very much different from interacting with students in the classroom. Given these conditions, there appears to be a great need for a valid research-based measure of the functional knowledge of teachers which would be relatively inexpensive, quick, and easy to administer and score. Additional efforts to simulate the decision-making process in the classroom more realistically than traditionally used paper-and-pencil tests would also be of great value.

One alternative that we have experienced some initial success with is a modification of traditionally used paper-and-pencil tests. This test, developed by Hays(1988), was the Simulation Test of Interactive Teaching Competencies-Hays (STITC-H).

The original STITC-H consisted of 65 problems depicting situations that occur in the actual classroom. These problems were developed by Hays and colleagues and then reviewed by a panel of teachers and administrators before the final selection of items was made (Hays, 1988). Each problem was presented in both a visual and auditory format using a slide projector and a

cassette player. After the description has been read, it remains visible while four possible alternative courses of action are presented, one at a time, on audiotape only. The examinee must decide whether each alternative is either appropriate or inappropriate given the situation described.

These alternatives are presented rapidly at intervals of five seconds. Each one is presented only once and must be judged independently of the others. For any given situation described, all, some, or none of the alternatives may be appropriate. Five seconds after the last alternative, a new problem appears on the screen followed by a series of alternatives presented orally.

The specifications for the Simulation Test of Interactive Teaching Competencies-Hays (STITC-H) were developed using a group of indicators of teacher competence specified by the Virginia Beginning Teacher Assistance Program (BTAP) (Bureau of Educational Research, 1984; Hays, 1988). The competencies included in the BTAP program were derived from a review of the literature on teacher effectiveness. This review consisted of a meta-analysis of five reviews of the research literature on effective teaching (Brophy & Good, 1986; Crawford & Robinson, 1983; Florida State Department of Education, 1980; Smith, 1983; Weber & Roff, 1983).

Preliminary Findings

Hays (1988) found that experienced teachers (n=46) performed significantly higher than student-teachers (n=30) in five of the

thirteen competencies as well as total test score. Both of these groups (n=76) who were trained in teacher education programs scored significantly higher than students without training in education (n=31) on the total test and nine of the thirteen competencies. These findings suggest that what this test measures is something which teachers learn from their training or from their experience as teachers. Also, because of the test's research base, it may be seen as an indication of the degree to which teachers apply research findings to teaching problems under the conditions simulating interactive decision-making found in the STITC-H.

Shannon (1990) revised the STITC-H, used by Hays (1988), and extended the findings found in the pilot study. Three specific contrasts were made to investigate the influence of professional training, classroom experience, and teaching experience. It was revealed that: 1) a group of students who had nearly completed their professional training in education (n=25) scored significantly higher than those students from the same institution without such training (n=78) on four competencies as well as the total test, 2) a group with classroom teaching experience (n=66) scored significantly higher than a group with no classroom experience (n=103) on four competencies as well as total test, and 3) there were significant differences between experienced classroom teachers (n=28) and student teachers (n=38) on four competencies with student teachers scoring higher on

three of them.

Summary

Preliminary evidence has been gathered in support of an alternative approach to testing teachers. The Simulation Test of Interactive Teaching Competencies-Hays (STITC-H) and its' revised form, the Simulation Exercise in Classroom Decision-Making, have been developed as research-based measures of the functional knowledge of teachers. Both the initial test construction and the revisions were carried out in accordance with a specified set of indicators of successful teaching outlined by the Beginning Teachers Assistance Program (BTAP). Both tests have also demonstrated initial success in producing evidence of construct validity.

Investigation of the predictive validity of this assessment strategy is currently underway. If this simulation exercise can be used to predict teacher performance, it could then be used as a screening device within both teacher education programs and in local school divisions. This exercise could also be used to identify strengths and weaknesses of practicing teachers so that the appropriate inservice training could be scheduled and other remedial services provided. An assessment strategy that measures this knowledge and ability to apply research findings to classroom situations would be of great value for both teacher evaluation and program evaluation purposes.

References

Association of Teacher Educators (1988). Teacher Assessment. Reston, VA: Author.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), Handbook o teaching, third edition, New York: Macmillan.

Bureau of Educational Research (1984). Beginning Teacher Assistance Program Phase II Final Report. Charlottesville: Author.

Chiu, Y. C. (1989). An Investigation of the Relationship between the National Teacher Examinations and a Performanced-Based Assessment of Professional Knowledge. Unpublished doctoral dissertation, University of Virginia, Charlottesville.

Crawford, J., & Robinson, C. (1983, March). A review of empirical research in classroom management. Classroom management: Reviews of the teacher education and research literature (pp. 43-62). Princeton, NJ: Educational Testing Service.

Eissenberg, T. E. & Rudner, L. M. (1988). State Testing of Teachers: A Summary of Current Practices. Washington, DC: American Institute for Research.

Florida State Department of Education (1980). Handbook of the Florida performance measurement system. Tallahassee, FL: Author.

Givens, S. (1988). ETS Unveils Plan to Replace National Teacher Examination. AACTE Briefs, 9(9-10), 11,16.

Hays, L. (1988). A Simulation Test of Teacher Competencies. Unpublished doctoral dissertation, University of Virginia,Carlottesville.

Hodgkins, R., & McKenna, B. (1982). Testing and teacher certification: An explosive combination. Educational Measurement: Issues and Practice, 1 (2). 10-26.

Medley, D. M., Rosenblum, E. P., & Vance, N. C. (1989). Assessing the Functional Knowledge of Participants in the Beginning Teacher Assistance Program. The Elementary School Journal, 89(4), 495-510.

Medley, D. M., Coker, H., & Soar, R. S. (1984). Measurement-based evaluation of teacher performance: An empirical approach. New York: Longman.

Medley, D. M. (1982). Teacher Competency Testing and the Teacher Educator. Charlottesville, VA: Association of Teacher Educators and the Bureau of Educational Research.

Medley, D. M. (1978). Some Obstacles to the Valid Measurement of Teacher Competence. A Paper presented at the Conference on Issues in Competency-Based Education, Georgia State University.

Piper, M. K., & Houston, W. W. (1981). The search for teacher competence: CBTE and MCT. Journal of Teacher Education, 31(5), 37-40.

Pottinger, P. (1979). Competence assessment: Comments on current practices. In P. S. Pottinger & J. Goldsmith (Eds.), Defining and measuring competence (25-39). San Francisco: Jossey-Bass.

Quirk, T., Witten, B., & Weinberg, S. (1973). Review of Studies of the Concurrent and Predictive Validity of the National Teacher Examinations. Review of Educational Research, 43 (1), 89-114.

Sandefur, J.T. (1985). Competency assessment of Teachers. Action in Teacher Education, 7(1-2), 1-6.

Scriven, M. (1981). Summative teacher evaluation. In J. Millman (Ed.). Handbook of teacher evaluation (244-271). Beverly Hills, CA: Sage.

Shannon, D. M. (1990). The Effects of Teacher Training and Experience in a Classroom Decision-Making Simulation. Unpublished doctoral dissertation, University of Virginia, Charlottesville.

Smith, D. C. (Ed.). (1983). Essential knowledge for beginning educators. Washington, DC: American Association of Colleges for Teacher Education and ERIC Clearinghouse on Teacher Education.

Smith, B. O. (1983). Teacher Education in Transition, in Essential knowledge for beginning educators (pp. 140-145mith (ed.). Washington, DC: American Association of Colleges for Teacher Education and ERIC Clearingho on Teacher Education.

Weber, W. A. & Roff, L. A. (1983). A review of the teacher education literature on classroom management. Classroom management: Reviews of the teacher education and research literature (pp. 7-42). Princeton, NJ: Educational Testing Service.